State of Vermont

Socrata Platform Pilot Testing
Geospatial Data

In Support of the VT GIS Data Warehouse
Redesign & Rebuild Project

Prepared By: Steve Sharp, VT Center for Geographic
Information
Date of Publication: 12/9/2013 (DRAFT v0.1)

Socrata Platform Pilot Testing – Geospatial Data
In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

Table of Contents

Revision History

Version	Date	Author(s)	Revision Notes
0.1	12/3/2013	Steve Sharp, VCGI	1st draft

_____

Overview - Platform Pilot Testing

VCGI has worked with DII to test Socrata's Open Data Portal product. VCGI conducted this testing in support of the VT GIS Data Warehouse Redesign/Rebuild project[1]. The Socrata pilot test was done in accordance with the Platform Pilot Testing Plan[2] crafted by the EGC Data Warehouse Workgroup.

Objectives

- Test the degree to which the Socrata Open Data platform supports the requirements[3] defined for the new VT GIS Data Warehouse.

Scope

In Scope:

- Test the Socrata Open Data platform against the requirements (user/publisher stories) identified for the VT GIS Data Warehouse.

Out of Scope:

- Loading all current VT GIS Data Warehouse entries into the Socrata Open Data platform.

Test Datasets

- **Crash data:** VTrans High crash locations (2006-2010).

- **Traffic data:** VTrans Average Annual Daily Traffic (2012).

- **Pavement condition:** VTrans pavement condition data (2012).

- **Broadband data:** VCGI broadband data (6/30/2013)

_____

[1] https://outside.vermont.gov/sites/egcgeo/dww/Shared%20Documents/VT-Data-Warehouse-Project-Charter-final-v0.6.pdf
[2] https://outside.vermont.gov/sites/egcgeo/dww/Shared%20Documents/VT-Data-Warehouse-PilotTestingPlan-final-v0.2.pdf
[3] https://outside.vermont.gov/sites/egcgeo/dww/Shared%20Documents/VT_DataWarehouse_UserStories_final.docx

_____

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild
_____

- **Aerial Photos:** VT orthophotos (4-band color imagery) - 2012

Procedures

- **Data Preparation:** Converted data to ESRI shapefile format (VT State Plane Meters NAD83). Followed VGIS Data Warehouse naming conventions.

- **Metadata Preparation:** Prepared FGDC CSDGM[4] compliant metadata for every dataset. Exported to ascii, XML, and HTML formats. Also exported to ISO 19139 XML format.

- **Data Loading and Tagging:** Zipped shapefiles + metadata and uploaded to VT Socrata site. Used the "Import Geospatial Data" for all except for VT orthos (used "Non-Data File" and "Linked to External Data" options). Manually entered basic information into Socrata's "metadata" form, and added keywords and tags. Made the data publically available (for those with access to the VT Socrata site).

- **Capabilities Review:** The goal of this pilot test was to evaluate the degree to which Socrata's Open Data platform supports the requirements[5] defined for the new VT GIS Data Warehouse. The results are documented in Attachment B.

Findings

Socrata's Open Data Portal has a rich set of tools used to load, publish, share, and use data. Its especially good at handling tablular datasets such as spreadsheets, with the ability to create multiple "filtered views" of the data along with graphs, and the option to download in many formats. The Open Data Portal can also consume geospatial datasets (eg: shapefiles), however, several limitations reduce the usability (for administrators and users) of these types of datasets within the platform. Attachment B includes a detailed review, but some important limitations have been highlight here.

- **Inability to change map symbology for shapefiles:** Raw geospatial data (map data) rendered with the same default color (in this case blue for lines and grey/black for polygons) has little to no meaning. Cartographic symbology adds meaning to a map. For example, the Average Annual Daily Traffic data's symbology should be modified so that areas with higher traffic volumes have a thicker lines and/or different colors vs areas with low traffic volumes. Unfortunately the Socrata platform doesn't

---

[4] http://www.fgdc.gov/metadata/geospatial-metadata-standards#csdgm
[5] https://outside.vermont.gov/sites/egcgeo/dww/Shared%20Documents/VT_DataWarehouse_UserStories_final.docx

_____

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild
_____

provide tools to change the symbology of imported shapefiles. Therefore the Traffic map is meaningless for anyone looking at it. The same is true for the broadband, pavement condition, and high crash location data.

- **Inability to apply filters to shapfiles to create different views:** The Socrata platform doesn't support filters with shapefiles. This impacts things such as the broadband dataset which includes 8 geodata files (eg: Cable layer, DSL layer, Fiber layer, etc..). Separate views need to be created for the map/data to have meaning to the general public.

- **250MB upload limit:** Files greater than 250MB can't be uploaded into Socrata's platform. The usual workaround (use Update Wizard "append" option) isn't supported with geospatial data (eg:shapefiles). Geospatial datasets often exceed the 250MB limit.

- **Imagery:** Socrata's platform has several limitations when it comes to raster imagery. The "Connect with an ESRI Map Layer" option does not support the use of ArcGIS Server "image services" (limited to "map services"). Therefore imagery services can't be visualized. Geo-referenced image files (rasters such as GeoTIFFs) can only be uploaded using the "Non-Data file" category, and therefore cannot be visualized. Also the 250MB size limit represents a serious limitations when it comes to imagery.

- **Clip & Zip not supported:**   Socrata's platform does not allow the user to select an "area of interest" for download. The user must download the entire dataset.

- **Metadata import not supported:** Socrata's platform cannot parse (harvest) metadata associated with a dataset including FGDC CSGDM and ISO 19139.

Conclusions

TDB

_____

Attachment A – Data Loading Log

Summary

Dataset	Prepared Shapefiles?	Prepared Metadata?	Zipped?	Uploaded?	Views?	Notes
TransRoad_PAVCON2012	Yes	Yes	Yes	Yes	No	Unable to modify look/feel of symbology on map. Unable to filter by layer name to create separate views for each layer
TransStats_AADT2012	Yes	Yes	Yes	Yes	No	Unable to modify look/feel of symbology on map
TransStats_HCL20062010	Yes	Yes	Yes	Yes	No	Unable to modify look/feel of symbology on map. Unable to filter by layer name to create separate views for each layer
UtilityTelecom_VTBB201306	Yes	Yes	Yes	Yes	No	Ran into 250MB load limit. Unable to modify look/feel of symbology on map. Unable to filter by layer name to create separate views for each layer
VT Orthophotos (2012)	NA	Yes	NA	Yes	No	Only able to uploaded 1 PAN image due to size limitations (TIFF format…not JP2). Testing create a "Link to External" dataset and pointed to WAF containing JP2s.

Detailed Step-by-Step

11/22/2013
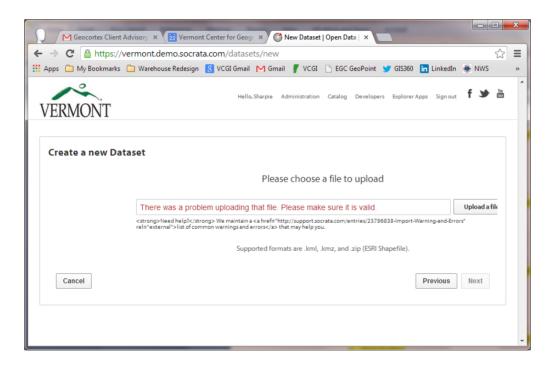
- Prepared AADT2012, PAVCON2012, HCL20062010, and VTBB201306 per VCGI standards.
    - Loaded into Feature Datasets, created FGDC metadata, loaded metadata into Feature Datasets
    - Created FGDC compliant metadata for all data. Exported to TXT, HTML, and XML format. Exported to ISO 19139 metadata format too (XML).
    - Exported data to shapefile format in GCS_WGS_1984 coordinate system and VT State Plane Coordinate System.
    - Zipped shapefiles, SP and WGS versions. WGS will be uploaded. SP will be attachment.

_____

- Uploaded to State of Vermont's Socrata site. Used "Import Geospatial" option
  - **TransStats_AADT2010.zip** (3,788KB)
    - Initiated upload. Detected shapefile and coordinate system.
    - Filled out descriptive info form.
      - Inserted ISO "transportation" keyword along with other keywords
    - Attachments
      - Zipped shapefiles (VT State Plane Coordinate System) + metadata
      - FGDC metadata in HTML format.
      - ISO 19139 metadata in XML format.
    - Unable to customize symbology of map ("Mondara layers are currently not customizable further.")
  - **TransRoad_PAVCON2012.zip** (9,389KB)
    - Initiated upload. Detected 2 shapefiles and coordinate system.
    - Filled out descriptive info form.
      - Inserted ISO "transportation" keyword along with other keywords
    - Attachments
      - Zipped shapefiles (VT State Plane Coordinate System) + metadata
      - FGDC metadata in HTML format.
      - ISO 19139 metadata in XML format.
    - Unable to customize symbology of map ("Mondara layers are currently not customizable further.") Both are rendering on top of each other.
  - **TransStats_HCL20062010.zip** (313KB)
    - Initiated upload. Detected 2 shapefiles and coordinate system.
    - Filled out descriptive info form.
      - Inserted ISO "transportation" keyword along with other keywords
    - Attachments
      - Zipped shapefiles (VT State Plane Coordinate System) + metadata
      - FGDC metadata in HTML format.
      - ISO 19139 metadata in XML format.

- - ▪ Unable to customize symbology of map ("Mondara layers are currently not customizable further.") Both shapefiles are rendering on top of each other.
  - ○ **UtilityTelecom_VTBB201306.zip** (395,259KB)
    - ▪ Initiated upload. Spun for a long time. Returned this error. Probably can't load a file this large.
      - • **Email from Darrell Cabales:** The 250 Mb upload limit is a soft limit. This is the threshold around which users sometimes experience timeout or system hangups. You can try uploading larger files, but just be on the lookout for inconsistent behavior. One common workaround is to upload the file in smaller chunks.
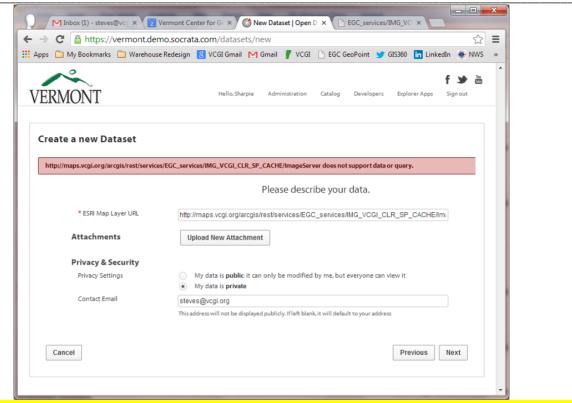
- • Test load of HCL20062010 shapefile in VT State Plane Meters instead of GCS WGS84. Realized that data doesn't need to be in GCS WGS84. Used the Edit → Replace Wizard to reload all shapefiles (HCL, AADT, PAVCON) in VT State Plane Meters. Dropped "Shapefiles in VT State Plane Meters" attachments from all datasets since its redundant.
- • Emailed LG regarding question about size limit for upload (250MB?).

_____

12/2/2013

- **UtilityTelecom_VTBB201306.zip** (162,312KB): I didn't get a reply from LG regarding size limit, so removed Utility_VTBB201306_poly_BlockAll shapefile from zip to reduce size of zip below 250MB. Added Utility_VTBB201306_poly_BlockPls instead.
  - Initiated upload. Detected 8 shapefiles and coordinate system. Took 15 minutes to load.
  - Filled out descriptive info form.
    - Inserted ISO "utilitiesCommunication" keyword along with other keywords
  - Attachments
    - FGDC metadata in HTML format.
    - ISO 19139 metadata in XML format.
  - Unable to customize symbology of map ("Mondara layers are currently not customizable further.") Shapefiles are rendering on top of each other.
- **2012 VT Orthophotos**
  - Tried to register dataset as "Connect with an ESRI Map Layer" by referencing this ArcGIS Server service (http://maps.vcgi.org/arcgis/rest/services/EGC_services/IMG_VCGI_CLR_WM_CACHE/ImageServer). It failed because it seems it can't use Image Services, only Map Services.

_____

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

o   Tried uploaded a JPEG2000 version of the file (1 ortho tile only) as a regular dataset, but this format is not support.

o   Tried uploaded a JPEG2000 version as a non-data file and the "Analyzing your file…" spun for a long time. Cancelled operation. Seems it can't handle JPEG2000 files.

o   Tried uploading a TIFF (GeoTIFF) as a non-data file, but the color orthophoto exceed 250MB so it was too big.

o   Uploaded a single black & white (panchromatic) tile in TIFF (GeoTIFF) as a non-data file.

-   Filled out descriptive info form.
    -   Inserted ISO "ImageryBaseMapsEarthCover" keyword along with other keywords
-   Attachments
    -   FGDC metadata in HTML format.
    -   Skipped ISO.

_____

- Unable to preview data, only able to download.
  - Registered "Link to External Data" and linked to WAF containing all 2012 color orthophotos (http://maps.vcgi.org/gisdata/vcgi/imagery/VTORTHO/0_5M/CLRIR/2012/COMP/)
    - Filled out descriptive info form.
      - Inserted ISO "ImageryBaseMapsEarthCover" keyword along with other keywords
    - Attachments
      - FGDC metadata in HTML format.
      - Skipped ISO.
    - Unable to preview data, only able to download.
- Filled out DII_Pilot_DataLoad_log.xlsx spreadsheet.
- Filled out Warehouse_PilotTestingReview_draft_v1.xlsx for Socrata.

_____

Attachment B – Detailed Results of Socrata Testing

User Stories/Requirements – Results of Socrata Pilot Test

Story - Requirement	Socrata Pilot Test	Socrata Pilot Test notes
As a User, I want "authoritative" datasets to be discoverable and clearly identified.	Not Tested	Maybe via keywords?
As a User, I want a browser-based Data Warehouse solution that does not require a browser plug-in.	Fully Supported	
As a User, I want a robust and easy-to-use search engine to help me find the geospatial data, web services, and web/mobile/desktop applications.	Partially Supported	Pretty good search engine. Fast. However, test search for the keyword "telecom" did not return Broadband dataset, but search for "telecommunications" did. Strange.
As a User, I want access to aerial photos (eg: VT orthos) with no downloading required (ie: as a web service).	Not Supported	Unable to use ArcGIS Server Image Services. Limited to Map Services only
As a User, I want access to historical aerial photos.	Fully Supported	
As a User, I want access to imagery organized by year (whether for download or web service).	Fully Supported	
As a User, I want access to VT GIS data layers with no downloading required (ie: as a web service).	Not Supported	SODA is so limited in this area that its basically useless
As a User, I want data and imagery downloads to include embedded metadata.	Workaround Possible	If imagery is zipped and metadata is included in the zip
As a User, I want metadata that is easy to read with a quick way to get to attribute descriptions and attribute codes.	Not Supported	Unable to harvest FGDC or ISO metadata

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

As a User, I want to download aerial photos (eg: VT orthos) so I can use them on my computer.	Workaround Possible	Unable to upload imagery due to file format and size limitations. Able to "link to external" as a workaround
As a User, I want to download VT GIS data layers and imagery (aerial photos) only for a particular area that I specify (eg: clip & zip).	Not Supported	
As a User, I want to download VT GIS data layers in ESRI file geodatabase format.	Workaround Possible	Possible to uploaded "attachment" as zipped FGDB.
As a User, I want to download VT GIS data layers in ESRI shapefile format.	Fully Supported	
As a User, I want access to "help" resources and media (such as YouTube videos) on how to use GIS data, web services, web applications, products (eg: layer package), and tools (eg: scripts and geoprocessing tools).	Not Tested	
As a User, I want access to data and web services which are openly available in a manner consistent with OpenDefinition.org's definition .	Not Tested	
As a User, I want access to data standards and data templates for download.	Fully Supported	Most standard document types can be posted.
As a User, I want access to ESRI layerfile packages.	Workaround Possible	Possible to uploaded "attachment" as zipped LPK files.
As a User, I want access to ESRI layerfiles (symbology) used to render GIS data layers.	Workaround Possible	Possible to uploaded "attachment" as zipped LPK files.
As a User, I want access to historical "snapshots" of VT GIS data.	Workaround Possible	Requires posting separate datasets for each "snapshot"
As a User, I want all data that I find to be available for download (not just a link to an external website).	Partially Supported	Some data can't be uploaded due to size or file format and therefore a link must be used

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

As a User, I want all of the data to have the good metadata and keyword tagging.	Partially Supported	Able to enter some metadata and tags. Unable to harvest FGDC or ISO metadata. Can only upload as "attachment"
As a User, I want more format/projection options when downloading.	Workaround Possible	Requires posting separate datasets or "attachments" for format/projection combo
As a User, I want the "posted date" included on the list of new/updated data, web services, or web/mobile/desktop applications.	Partially Supported	Results can be sorted by newest/oldest
As a User, I want to be notified whenever new or updated data, web services, or web/mobile/desktop applications are released.	Not Tested	
As a User, I want to download all data as pre-packaged files via FTP.	Not Supported	Socrata doesn't offer an FTP portal. Would require posting of separate data to an FTP site.
As a User, I want to download data in non-GIS formats (eg: CSV, GeoJSON).	Partially Supported	Yes many non-GIS formats supported. Not GeoJSON. KML/KMZ supported but it's a "raster overlay" instead of vector.
As a User, I want to download thematic data in bulk (eg: all transportation data).	Not Supported	Would require bundling of data by theme, however, this would most likely exceed the 250MB upload limit.
As a User, I want to download VT GIS data layers in KML format.	Partially Supported	Yes but KML/KMZ supported but it's a "raster overlay" instead of vector. Not very usable.
As a User, I want to explore Warehouse entries by the "custom data groupings" created by the Publisher.	Fully Supported	Done via keywords
As a User, I want to find and download Regional-level (RPC) GIS data.	Fully Supported	However, RPCs may not have access to the systems.

_____

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

As a User, I want to find and download town-level GIS data.	Fully Supported	However, towns may not have access to the system
As a User, I want to find the latest GIS data releases by agency, by department, by program (eg: permitting/regulatory/monitoring).	Not Tested	
As a User, I want to flag errors in the data and have them submitted to the data steward.	Not Tested	
As a User, I want to get to all Warehouse entries that have one or more "custom keyword tags" assigned by the Publisher.	Fully Supported	
As a User, I want to have access to Data Warehouse "capabilities" via a published API (eg: REST endpoint) so I can integrate it into my own website.	Not Tested	Maybe via SODA
As a User, I want to integrate aerial photos (eg: VT orthos) into my own web mashups.	Not Supported	Would use ArcGIS Server for this
As a User, I want to integrate VT GIS data layers into my own web mashups.	Partially Supported	Very limited via Socrata Open API (SODA).
As a User, I want to post comments about a particular dataset, web service, or application and have those comments show up whenever someone discovers the dataset or web service.	Fully Supported	
As a User, I want to search by standard data themes (eg: boundary data, transportation data, etc.) using ISO 19115 Topic Categories standard.	Not Tested	Administrator can probably create custom categories such as these
As a User, I want to see a listing of all available data even if it isn't available for immediate download or if there are access constraints.	Fully Supported	
As a User, I want web services providing access to aerial photos that render "Google" fast.	Not Supported	
As a User, I want web services that work with ArcGIS software (desktop, mobile, and browser-based clients).	Not Supported	
As a User, I want web services that work with Open Source GIS software (desktop, mobile, and browser-based clients).	Not Supported	

_____

_____

As a User, I want a button that allows me to "share" things that I discover on the Data Warehouse via a URL link, Twitter and Facebook.	Fully Supported	
As a User, I want all technical terms to be linked to explanations so I don't have to be a GIS expert.	Not Tested	
As a User, I want to browse thumbnails of available datasets, web services, or web application to give me a quick sense of what is being offered.	Fully Supported	
As a User, I want to control layer rendering (symbology) for VT GIS data layers accessed via a web map service.	Not Supported	
As a User, I want to control when and what I get notified about.	Not Tested	
As a User, I want to download VT GIS data layers in GML format.	Not Supported	
As a User, I want to personalize the site (eg: My Warehouse) for the topics and data that I use most often.	Not Supported	
As a User, I want to post my "data wish list" to a website for others to see, rank, and comment on.	Not Tested	
As a User, I want to rank (eg: number of stars) datasets, web services, and web applications.	Fully Supported	
As a User, I want to see a list of all data, web services, and applications associated with a particular dataset.	Not Tested	
As a User, I want to see rankings of datasets, web services, and web applications to help me pick the "best" option.	Fully Supported	
As a User, I want to specify data types (eg: polygon, points, lines, raster) in searches.	Not Supported	

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

Publisher Stories/Requirements – Results of Socrata Pilot Test

Story - Requirement	Socrata Pilot Test	Socrata Pilot Test notes
As a Publisher, I want a publishing solution that requires a minimal number of manual steps.	Unknown	Can't really determine this since Socrata doesn't meet all of the requirements and therefore would need to be augmented with other solutions which might results in many manual steps.
As a Publisher, I want GIS customers to have access to my web mapping applications from the data warehouse.	Workaround Possible	
As a Publisher, I want my data to be discoverable and linked through other catalogs (eg: data.gov, ArcGIS Online, etc.) so that it can be found by users outside of Vermont.	Not Tested	I don't believe there is any way to link directly to data.gov. Would need to continue to maintain FGDC/ISO metadata in a WAF for data.gov to harvest
As a Publisher, I want my metadata to conform to standards.	Not Supported	Cannot harvest ISO or FGDC metadata
As a Publisher, I want the ability to include non-geospatial data files (eg: MS Word document, Excel spreadsheet) with the geospatial data that I share.	Fully Supported	
As a Publisher, I want the search engine to be driven from the metadata that I submit.	Not Supported	Cannot harvest ISO or FGDC metadata

_____

As a Publisher, I want the standardized metadata to automatically populate specific attributes of the online catalog entries (website content).	Not Supported	Cannot harvest ISO or FGDC metadata
As a Publisher, I want the Warehouse to offer a simple data hosting subscription plan for non-governmental organizations and/or individuals.	Not Applicable	Would be hosted by VT DII. Those outside the state would not have access.
As a Publisher, I want to authenticate to my publishing account.	Fully Supported	
As a Publisher, I want to be able to define groups of users that can collaborate on editing and maintaining the data with me.	Not Tested	
As a Publisher, I want to be able to post data, web services, applications (web/mobile/desktop), products (eg: layer packages), and tools (eg: scripts and geoprocessing tools).	Workaround Possible	Data, tools, and "products" is generally supported with limitations. Web Applicatoins and Web services (except for AGS Map Services) are not other than "Link to External".
As a Publisher, I want to be in control of my own publishing, information management, and publisher account.	Partially Supported	Limited controls
As a Publisher, I want to have access to download statistics of my data so I know how often my data is downloaded.	Not Tested	
As a Publisher, I want to publish my data as services that work on multiple clients including desktop software, mobile devices, and web browsers.	Partially Supported	Very limited via SODA
As a Publisher, I want a Warehouse solution that complies with W3C DCAT and OGC CS specifications.	Not Tested	DCAT maybe? OGC CS not supported

Socrata Platform Pilot Testing – Geospatial Data

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

As a Publisher, I want a Warehouse solution that make it easy for me to comply with data archival procedures defined by the State of Vermont.	Not Tested	
As a Publisher, I want a Warehouse solution that is able to harvest ISO and FGDC compliant metadata records.	Not Supported	
As a Publisher, I want a mechanism for managing "data/file locks" when posting data updates.	Fully Supported	Socrata offers an update data option and/or ability to make data private.
As a Publisher, I want the ability to "share" datasets, web services, or web/mobile/desktop applications that I publish on the Data Warehouse via Twitter and Facebook.	Fully Supported	
As a Publisher, I want the ability to assign myself to specific "stakeholder groups" (eg: private sector, government, non-profit, etc.) so that users can look for Warehouse entries by these groups.	Not Tested	
As a Publisher, I want to be able to moderate comments about a particular dataset or web service I have published.	Not Tested	
As a Publisher, I want to be able to push out data, web service, or application update alerts to subscribers.	Not Tested	
As a Publisher, I want users to be notified with a special icon/flag when a particular dataset, web service, or web/mobile/desktop applications is/are temporarily unavailable.	Not Supported	Can only change to "private" so it isn't visible.
As a Publisher, I want a mechanism which allows me temporarily disable access to a particular dataset, web service, or web application.	Fully Supported	Can switch to "private" so it can't be seen
As a Publisher, I want my data, web services, and web applications to be discoverable through Google searches so that it is more likely to be found.	Not Tested	

_____

In Support of VT GIS Data Warehouse - Redesign & Rebuild

_____

As a Publisher, I want my own configurable "space" or "site" on the Data Warehouse with custom branding, content, and messaging.	Not Tested	
As a Publisher, I want the ability to add "custom keyword tags" to my Warehouse entries to help users find what they are looking for.	Fully Supported	
As a Publisher, I want the ability to create my own "custom data groupings".	Fully Supported	Done via keywords
As a Publisher, I want the ability to tag datasets, web services, and web/mobile/desktop applications to one or more ISO 19115 Topic Categories (eg: boundary data, transportation data, etc).	Not Tested	
As a Publisher, I want to clearly identify and "brand" data, web services, and applications provided by my organization.	Not Tested	
As a Publisher, I want to have access to statistics showing how often particular web services are used.	Not Tested	